

## SEMICONDUCTOR THEORY

### D-1 INTRODUCTION

Semiconductor devices are the fundamental building blocks from which all types of useful electronic products are constructed—amplifiers, high-frequency communications equipment, power supplies, computers, control systems, to name only a few. It is possible to learn how these devices—diodes, transistors, integrated circuits—can be connected together to create such useful products with little or no knowledge of the semiconductor theory that explains how the devices themselves operate. However, the person who understands that theory has a greater knowledge of the capabilities and limitations of devices and is therefore able to use them in more innovative and efficient ways than the person who does not. Furthermore, it is often the case that the success or failure of a complex electronic system can be traced to a certain peculiarity or operating characteristic of a single device, and an intimate knowledge of how and why that device behaves the way it does is the key to reliable designs or to practical remedies for substandard performance.

In Sections D-2 through D-7, we present the fundamental theory underlying the flow of charged particles through semiconductor material, the material from which all modern devices are constructed. We will also learn the theory of operation of the most fundamental semiconductor device: the diode. The theory presented here includes numerous equations that allow us to compute and assign quantitative values to important atomic-level properties of semiconductors.

### D-2 ATOMIC STRUCTURE

A study of modern electronic devices must begin with a study of the materials from which those devices are constructed. Knowledge of the principles of material composition, at the level of the fundamental structure of matter, is

an important prerequisite in the field of study we call *electronics* because the ultimate concern of that field is predicting and controlling the flow of atomic charge. Towards developing an appreciation of how atomic structure influences the electrical properties of materials, let us begin with a review of the structure of that most fundamental of all building blocks: the atom itself.

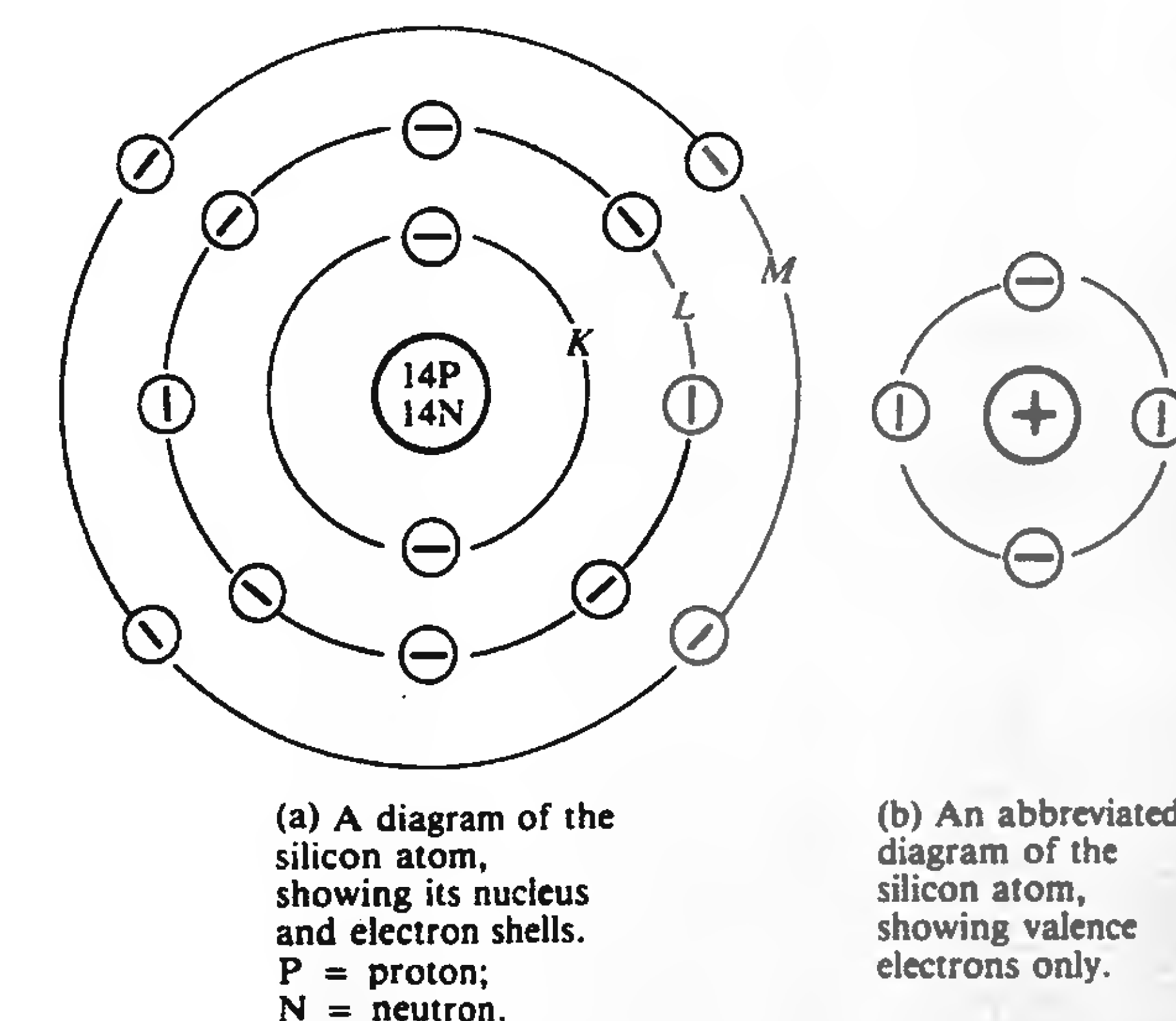
Every chemical element is composed of atoms, and all of the atoms within a *single* element have the same structure. Each element is unique because that common structure of its atoms is unique. Each atom is itself composed of a central *nucleus* containing one or more positively charged particles called *protons*. When an atom is complete, its nucleus is surrounded by negatively charged particles, called *electrons*, equal in number to the quantity of protons in the nucleus. Since the positive charge on each proton is equal in magnitude to the negative charge on each electron, the complete atom is electrically neutral. Depending on the element that a particular atom represents, the nucleus may also contain particles called *neutrons* that carry no electrical charge. An atom from the simplest of all elements, hydrogen, contains exactly one electron and has a nucleus composed of one proton and no neutrons. All other elements have at least one neutron.

Figure D-1(a) is a diagram of the structure of an isolated atom of the element *silicon*, the material used most often in the construction of modern electronic devices. Rather than showing a cluster of individual protons and neutrons in the nucleus, the figure simply shows that the nucleus contains 14 (positively charged) protons and 14 neutrons. Notice that the atom is neutral, since it contains a total of 14 (negatively charged) electrons. These electrons are shown distributed among three distinct *orbits* around the nucleus. The electrons within a given orbit are said to occupy an *electron shell*, and it is well known that each shell in an atom can contain no more than a certain maximum number of electrons. If the first four shells are numbered in sequence beginning from the innermost shell (shell number 1 being that closest to the nucleus), then the maximum number ( $N_e$ ) of electrons that shell number  $n$  can contain is

$$N_e = 2n^2 \quad (\text{D-1})$$

Notice in Figure D-1(a) that shell number 1 (labeled K) is filled because it contains  $2 \times 1^2 = 2$  electrons and shell number 2 (labeled L) is filled because

FIGURE D-1





it contains  $2 \times 2^2 = 8$  electrons. However, shell number 3 (*M*) is not filled because it has a capacity of  $2 \times 3^2 = 18$  electrons but contains only 4.

Each shell is divided into *subshells*, the *n*th shell having *n* subshells. The first subshell in an electron shell can contain 2 electrons. If the shell has a second subshell, it can contain 4 additional, or 6 total, electrons; and if there is a third subshell, it can contain four more, or 10, electrons. The first shell of silicon has only one subshell ( $n = 1$ ) and it is filled (2 electrons); the second shell has two ( $n = 2$ ) filled subshells (containing 2 and 6 electrons); and the third shell has one filled subshell containing 2 electrons. The second subshell of the third shell can contain 6 electrons but has only 2. In practice, shells are given letter designations rather than numerical ones, so shells 1, 2, 3, 4, . . . , are referred to as shells *K*, *L*, *M*, *N*, . . . , as shown in Figure D-1(a). Subshells are designated *s*, *p*, *d*, *f*.

## EXAMPLE D-1

The nucleus of a germanium atom has 32 protons. Assuming that each shell must be filled before a succeeding shell can contain any electrons, determine the number of electrons in each of its shells and in the subshells of each shell.

## Solution

There are 32 electrons, and they fill the shells as shown in the *contents* column of the following table:

Shell	Capacity ( $2n^2$ )	Contents
<i>K</i> ( $n = 1$ )	2	2
<i>L</i> ( $n = 2$ )	8	8
<i>M</i> ( $n = 3$ )	18	18
<i>N</i> ( $n = 4$ )	32	4
		Total: 32

The contents of the subshells in each shell are shown in the following table:

Shell	Subshells	Capacity	Contents
<i>K</i>	<i>s</i>	2	2
<i>L</i>	<i>s</i>	2	2
	<i>p</i>	6	6
<i>M</i>	<i>s</i>	2	2
	<i>p</i>	6	6
	<i>d</i>	10	10
<i>N</i>	<i>s</i>	2	2
	<i>p</i>	6	2
	<i>d</i>	10	0
	<i>f</i>	14	0
			Total: 32

Not every electron in every atom is constrained forever to occupy a certain shell or subshell of an atomic nucleus. Although electrons tend to remain in their shells because of their force of attraction to the positively charged nucleus, some of them acquire enough energy (as, for example, from heating) to break away from their "parent" atoms and wander randomly through the

material. Electrons that have escaped their shells are called *free* electrons. Conductors have a great many free electrons, while insulators have relatively few.

The outermost shell in an atom is called the *valence* shell, and the number of electrons in the valence shell has a significant influence on the electrical properties of an element. The reason that the number of valence electrons is important is that electrons in a nearly empty valence shell or subshell are more easily dislodged (freed) than electrons in a filled or nearly filled shell. Moreover, valence electrons, being farther from the nucleus than electrons in the inner shells, are the ones that experience the least force of attraction to the nucleus. Conductors are materials whose atoms have very few electrons in their valence shells (copper atoms have only one), and, in these materials, the heat energy available at room temperature (25°C) is enough to free large numbers of those loosely bound electrons. When an electrical potential is applied across the ends of a conductor, free electrons readily move from one end to the other, creating a transfer of charge through the conductor, i.e., an electrical current. Valence electrons in insulators, on the other hand, are tightly bound to their parent atoms.

Because we will be concerned primarily with the behavior of electrons in the valence shell, we will abbreviate all future diagrams of atoms so that only the nucleus and the valence electrons are shown. The silicon atom is shown in this abbreviated manner in Figure D-1(b).

## D-3 SEMICONDUCTOR MATERIALS

Virtually all modern electronic devices are constructed from *semiconductor* material. As the name implies, a semiconductor is neither an electrical insulator (like rubber or plastic) nor a good conductor of electric current (like copper or aluminum). Furthermore, the mechanism by which charge flows through a semiconductor cannot be entirely explained by the process known to cause the flow of charge through other materials. In other words, a semiconductor is something more than just a conductor that does not conduct very well or an insulator that allows some charge to pass through it.

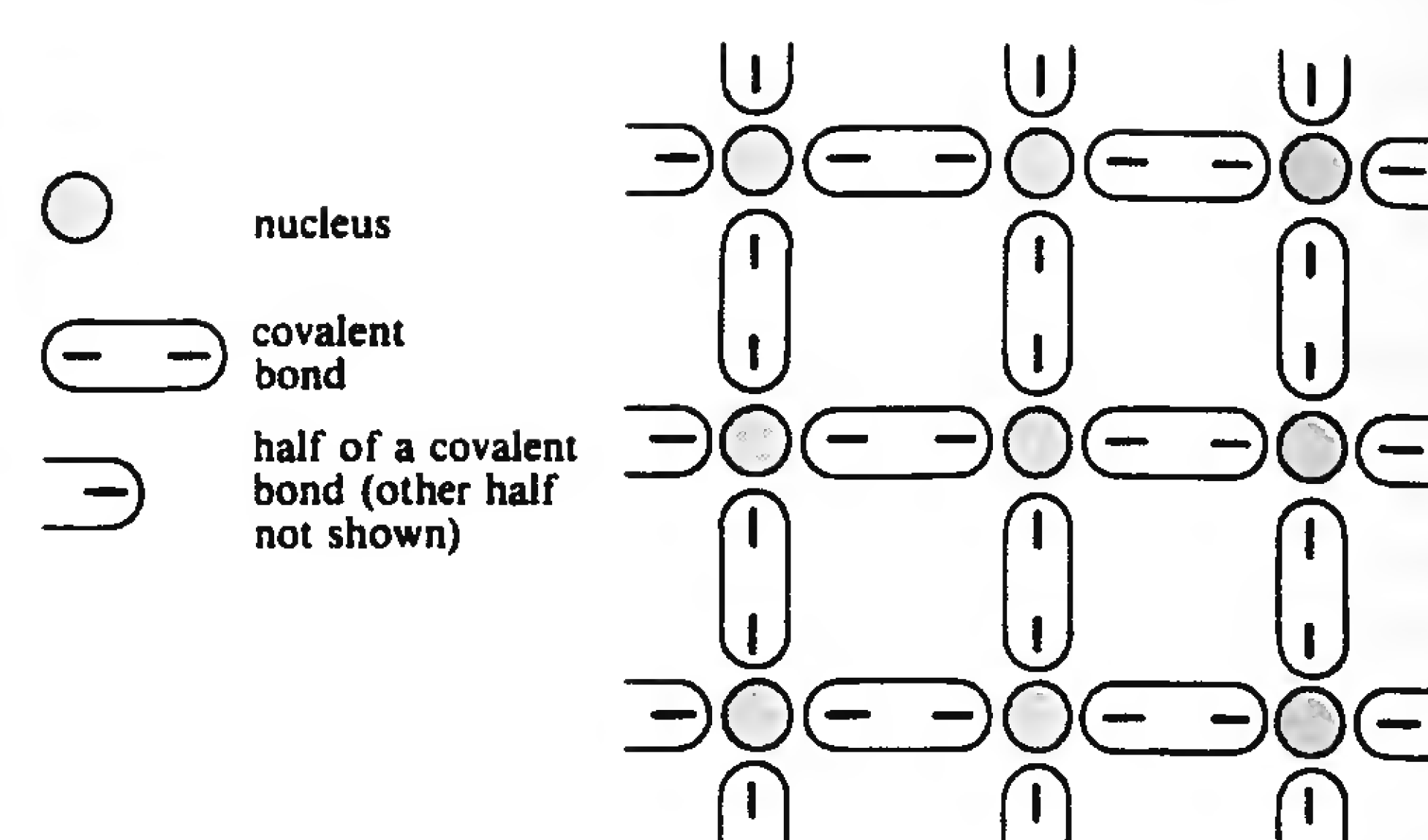
The electrical characteristics of a semiconductor stem from the way its atoms interlock with each other to form the structure of the material. Recall that conductors have nearly empty valence shells and tend to produce free electrons, while insulators tend to retain valence electrons. The valence shell of a semiconductor atom is such that it can just fill an incomplete subshell by acquiring four more electrons. For example, we have already noted that the *p* subshell of the *M* shell in silicon contains 2 electrons. Given four more, this subshell would be filled. A semiconductor atom seeks this state of stability and achieves it by *sharing* the valence electrons of four of its neighboring atoms. It in turn shares each of its own four electrons with its four neighbors and thus contributes to the filling of their subshells. Every atom duplicates this process, so every atom uses four of its own electrons and one each from four of its neighbors to fill its *p* subshell. The result is a stable, tightly bound, lattice structure called a *crystal*.

The interlocking of semiconductor atoms through electron sharing is called *covalent bonding*. It is important to be able to visualize this structure, and Figure D-2 shows a two-dimensional representation of it. Of course, a true crystal is a three-dimensional object—the covalent bonding occurs in all directions—but the figure should help clarify the concept of covalent bonding. Remember that we show only the valence electrons of each atom.

In Figure D-2, each pair of shared electrons forms a *covalent bond*, which is shown as two electrons enclosed by an oval. Note that only the center atom



FIGURE D-2 Covalent bonding in a semiconductor crystal



is shown with a complete set of 4 covalent bonds, but all of the other atoms would be similarly interlocked with their neighbors. As previously described, the center atom uses its own four electrons and one from each of four neighbors, so it effectively has 8 electrons in its  $M$  shell. The atom directly above the center one similarly has four of its own electrons and one from each of four neighbors (although only three are shown), so it too has 8.

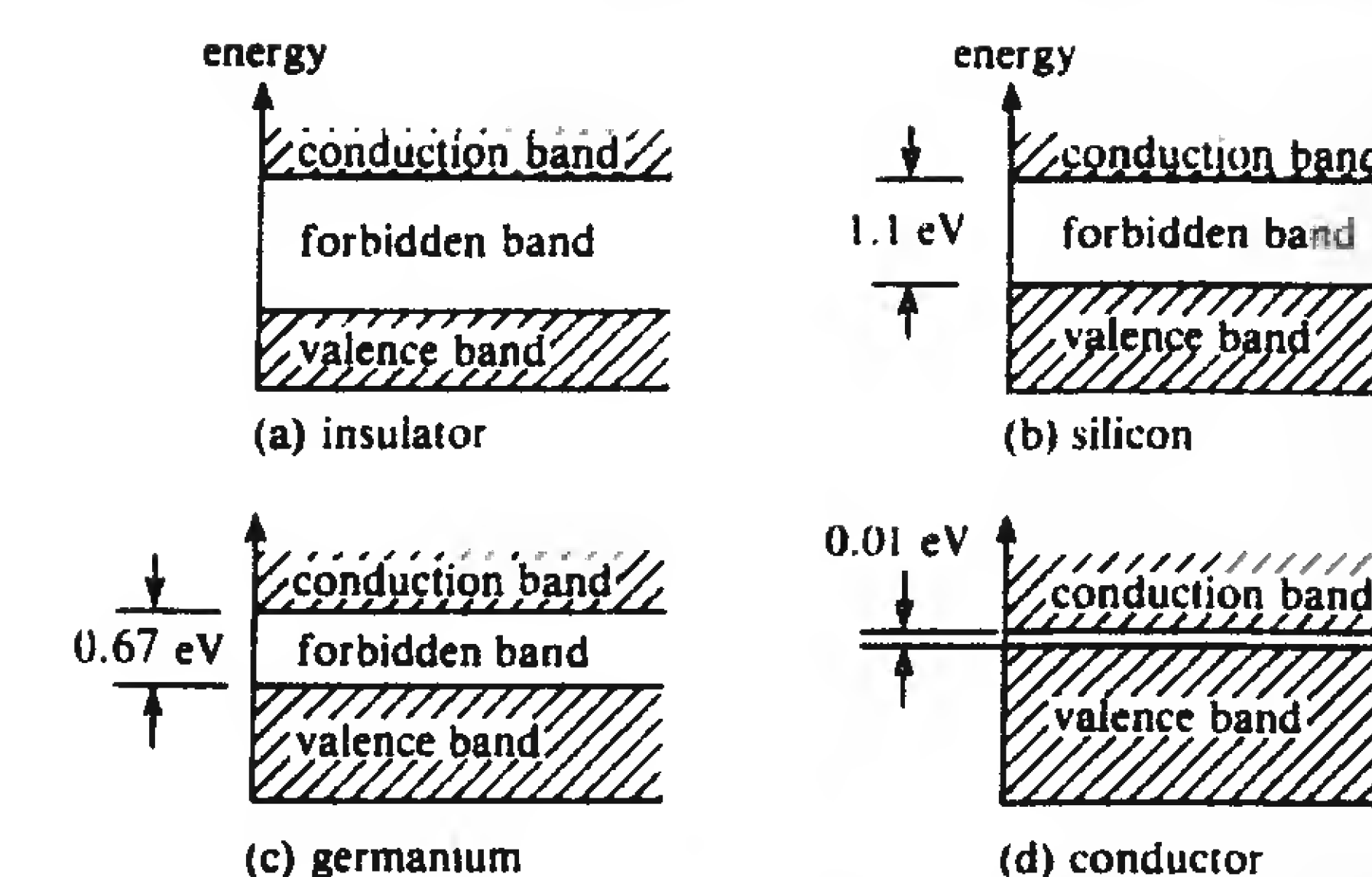
Germanium (symbol Ge) is another element whose valence shell enables it to establish covalent bonds with its neighbors and form a crystalline structure. In Example D-1, we saw that its  $N$  shell contains 4 electrons, 2 of which are in the  $p$  subshell. Therefore, it too can complete a subshell by the acquisition of 4 electrons. Like silicon (symbol Si), germanium atoms interlock and form a semiconductor material that is used to construct electronic devices. (However, because of temperature-related properties that we will discuss later, germanium is not now so widely used as silicon.) Note that the diagram in Figure D-2 applies equally to germanium and silicon, since only valence shell electrons are shown. Carbon also has a valence shell that enables it to establish covalent bonds and form a crystal, but it assumes that form only after being subjected to extreme heating and pressure. A carbon crystal is in fact a *diamond*, and it is not used in the construction of semiconductor devices.

#### D-4 CURRENT IN SEMICONDUCTORS

We have mentioned that the source of electrical charge available to establish current in a conductor is the large number of free electrons in the material. Recall that an electron is freed by acquiring energy, typically heat energy, that liberates it from a parent atom. Electrons are freed in semiconductor materials in the same way, but a greater amount of energy is required, on the average, because the electrons are held more tightly in covalent bonds. When enough energy is imparted to an electron to allow it to escape a bond, we say that a covalent bond has been *ruptured*.

It is instructive to view the electron-liberation process from the standpoint of the quantity of energy possessed by the electrons. The unit of energy that is conventionally used for this purpose is the *electron volt* (eV), which is the energy acquired by 1 electron if it is accelerated through a potential difference of 1 volt. One eV equals  $1.602 \times 10^{-19}$  joules (J). According to modern quantum theory, an electron in an isolated atom must acquire a very specific amount of energy in order to be freed, the amount depending on the kind of atom to which it belongs and the shell it occupies. Electrons in a valence shell already possess considerable energy, because a relatively small amount of additional energy will liberate them. Electrons in inner shells possess little energy, since they are strongly attracted to the nucleus and would

FIGURE D-3 Energy-band diagrams for several different materials



therefore need a great deal of additional energy to be freed. Electrons can also move from one shell to a more remote shell, provided they acquire the distinct amount of energy necessary to elevate them to the energy level represented by the new shell. Furthermore, electrons can *lose* energy, which is released in the form of heat or light, and thereby fall into lower-energy shells. Free electrons, too, can lose a specific amount of energy and fall back into a valence shell.

When atoms are in close proximity, as they are when they are interlocked to form a solid material, the interactions between adjoining atoms make the energy levels less distinct. In this case, it is possible to visualize a nearly continuous *energy band*, and we refer to electrons as occupying one band or another, depending on their roles in the structure. Energy-band diagrams are shown in Figure D-3. Free electrons are said to be in the *conduction band* because they are available as charge carriers for the conduction of current. *Valence-band* electrons have less energy and are shown lower in the energy diagram. The region between the valence and conduction bands is called a *forbidden band*, because quantum theory does not permit electrons to possess energies at those particular levels. The width of the forbidden band is the *energy gap* that electrons must surmount to make the transition from valence band to conduction band. Note in Figure D-3(a) that a typical insulator has a large forbidden band, meaning that valence electrons must acquire a great deal of energy to become available for conduction. For example, the energy gap for carbon is 5.4 eV. Energy gaps for semiconductors depend on temperature. As shown in Figure D-3, the room-temperature values for silicon and germanium are about 1.1 eV and 0.67 eV, respectively. The energy gap for a conductor is quite small ( $\leq 0.01$  eV) or nonexistent, and the conduction and valence bands are generally considered to overlap.

As might be surmised from the foregoing discussion, the number of free electrons in a material, and consequently its electrical *conductivity*, is heavily dependent on temperature. Higher temperatures mean more heat and therefore greater electron energies. At absolute zero ( $-273^\circ\text{C}$ , or 0 K), all electrons have zero energy. But as the temperature is raised, more and more electrons acquire sufficient energy to cross the gap into the conduction band. For a semiconductor, the result is that conductivity increases with temperature (resistance decreases), which means that a semiconductor has a *negative* temperature coefficient of resistance. Although the number of conduction-band electrons in a conductor also increases with temperature, there are so many more of these than in a semiconductor that another effect predominates: Their number becomes so vast that they collide frequently and interfere with each other's progress when under the influence of an applied electric potential. The increased heat energy imparted to them at higher



temperatures also makes their motion more erratic and compounds the problem. Consequently, it becomes more difficult to establish a uniform flow of charge at higher temperatures, resulting in a positive temperature coefficient of resistance for conductors.

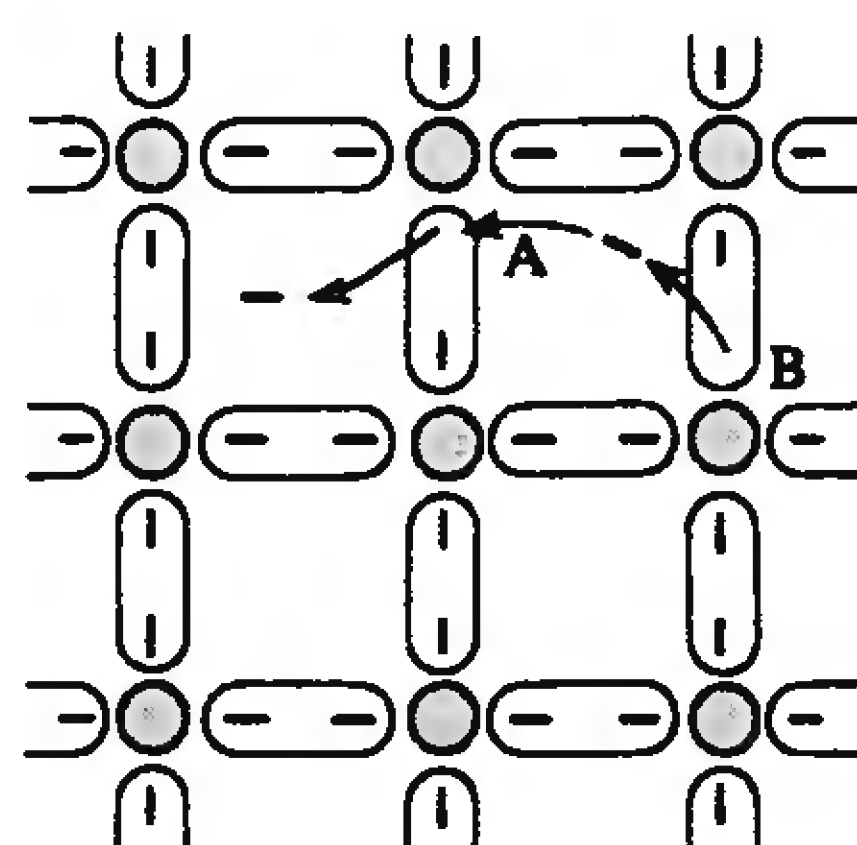
### Holes and Hole Current

What really distinguishes electrical current in a conductor from that in a semiconductor is the existence in the latter of another kind of charge flow. Whenever a covalent bond in a semiconductor is ruptured, a *hole* is left in the crystal structure by virtue of the loss of an electron. Since the atom that lost the electron now has a net positive charge (the atom becomes what is called a positive *ion*), we can regard that hole as representing a unit of positive charge. The increase in positive charge is, of course, equal to the decrease in negative charge, i.e., the charge of one electron:  $q_e = 1.6 \times 10^{-19}$  coulombs (C). If a nearby valence-band electron should now enter the hole, leaving behind a new hole, then the net effect is that a unit of positive charge has moved from the first atom to the second. This transfer of a hole from one atom to another constitutes a flow of (positive) charge and therefore represents a component of electric current, just as electron flow contributes to current by the transfer of negative charge. We can therefore speak of *hole current* in a semiconductor as well as electron current. Figure D-4 illustrates the concept.

Figure D-4 illustrates the *single* repositioning of a hole, but it is easy to visualize still another valence electron entering the new hole (at B in the figure), causing the hole to move again, and so forth, resulting in a hole path through the crystal. Note that holes moving from left to right cause charge transfer in the same direction as electrons moving from right to left. It would be possible to analyze semiconductor current as two components of electron transfer, but it is conventional to distinguish between conduction-band electron flow and valence-band hole flow. Here is an important point that is worth repeating because it is a source of confusion to students and is not emphasized enough in most textbooks: Hole current occurs at the *valence-band* level, because valence-band electrons do not become free electrons when they simply move from atom to atom. Electron current always occurs in the *conduction band*, and involves only the flow of free electrons. If a conduction-band electron falls into a hole (which it may), this does *not* constitute current flow; indeed, such an occurrence is a cancellation of charge, and we say that a hole-electron pair has been *annihilated*, or that a *recombination* has occurred. Since charge transfer can take place by the motion of either negatively charged electrons or positively charged holes, we refer to electrons and holes collectively as *charge carriers*. Note that hole current does not occur in a conductor.

Because holes in a (pure) semiconductor are created by electrons that have been freed from their covalent bonds, the number of free electrons must equal the number of holes. This equality applies to the semiconductor materials we have studied so far because we have assumed them to be pure (in the

FIGURE D-4 Hole current. When the electron in position A is freed, a hole is left in its place. If the electron in position B moves into the hole at A, the hole, in effect, moves from A to B.



sense that they are composed exclusively of atoms from one kind of element). Later, we will study semiconductor materials that have been made impure purposely in order to change the balance between holes and electrons. Pure semiconductor material is said to be *intrinsic*. It follows that the electron density, in electrons/m<sup>3</sup>, equals the hole density, holes/m<sup>3</sup>, in an intrinsic semiconductor. The subscript *i* is used to denote an intrinsic property;  $n_i$  refers to intrinsic electron density, and  $p_i$  is intrinsic hole density. Thus,

$$n_i = p_i \quad (\text{D-2})$$

At room temperature, the charge carrier densities for germanium and silicon are approximately  $n_i = p_i = 2.4 \times 10^{19}$  carriers/m<sup>3</sup> for germanium and  $n_i = p_i = 1.5 \times 10^{16}$  carriers/m<sup>3</sup> for silicon. These figures seem to imply vast numbers of carriers per cubic centimeter, but consider the fact that a cubic centimeter of silicon contains more than  $10^{22}$  atoms. Thus, there are approximately  $10^{12}$  times as many atoms as there are carriers in silicon. Consider also that the conductor copper contains approximately  $8.4 \times 10^{28}$  carriers (free electrons) per cubic meter, a carrier density that is immensely greater than that of either germanium or silicon. Note that the carrier density of germanium is greater than that of silicon because the energy gap, as shown in Figure D-3, is smaller for germanium than for silicon. At a given temperature, the number of germanium electrons able to escape their bonds and enter the conduction band is greater than the number of silicon electrons that can do likewise.

### Drift Current

When an electric potential is applied across a semiconductor, the electric field established in the material causes free electrons to drift in one direction and holes to drift in the other. Because the positive holes move in the opposite direction from the negative electrons, these two components of current *add* rather than cancel. The total current due to the electric field is called the *drift current*. Drift current depends, among other factors, on the ability of the charge carriers to move through the semiconductor, which in turn depends on the type of carrier and the kind of material. The measure of this ability to move is called *drift mobility* and has the symbol  $\mu$ . Following are typical values for hole mobility ( $\mu_p$ ) and electron mobility ( $\mu_n$ ) in germanium and silicon:

Silicon	Germanium
$\mu_n = 0.14 \text{ m}^2/(\text{V} \cdot \text{s})$	$\mu_n = 0.38 \text{ m}^2/(\text{V} \cdot \text{s})$
$\mu_p = 0.05 \text{ m}^2/(\text{V} \cdot \text{s})$	$\mu_p = 0.18 \text{ m}^2/(\text{V} \cdot \text{s})$

Note that the units of  $\mu$  are square meters per volt-second. Recall that the units of electric field intensity are V/m, so  $\mu$  measures carrier velocity (m/s) per unit field intensity:  $(\text{m/s})/(\text{V/m}) = \text{m}^2/(\text{V} \cdot \text{s})$ . It follows that

$$v_n = \bar{E}\mu_n \text{ and } v_p = \bar{E}\mu_p \quad (\text{D-3})$$

where  $\bar{E}$  is the electric field intensity in V/m and  $v_n$  and  $v_p$  are the electron and hole velocities in m/s. Although the value of  $\mu$  depends on temperature and the actual value of the electric field intensity, the values listed above are representative of actual values at low to moderate field intensities and at room temperature.

We can use carrier mobility to compute the total *current density*  $J$  in a semiconductor when the electric field intensity is known. Current density is current per unit cross-sectional area.



$$J = J_n + J_p = nq_n\mu_n\bar{E} + pq_p\mu_p\bar{E} \\ = nq_nv_n + pq_p v_p \quad (\text{D-4})$$

where  $J$  = current density, A/m<sup>2</sup>

$n, p$  = electron and hole densities, carriers/m<sup>3</sup>  
 $q_n = q_p$  = unit electron charge =  $1.6 \times 10^{-19}$  C  
 $\mu_n, \mu_p$  = electron and hole mobilities, m<sup>2</sup>/(V · s)  
 $\bar{E}$  = electric field intensity, V/m  
 $v_n, v_p$  = electron and hole velocities, m/s

Equation D-4 expresses the fact that the total current density is the sum of the electron and hole components of current density,  $J_n$  and  $J_p$ . For intrinsic material, equation D-4 can be simplified as follows:

$$J = n_i q_n \bar{E} (\mu_n + \mu_p) = p_i q_p \bar{E} (\mu_n + \mu_p) \\ = n_i q_n (v_n + v_p) = p_i q_p (v_n + v_p)$$

An analysis of the units of equation D-4 shows that

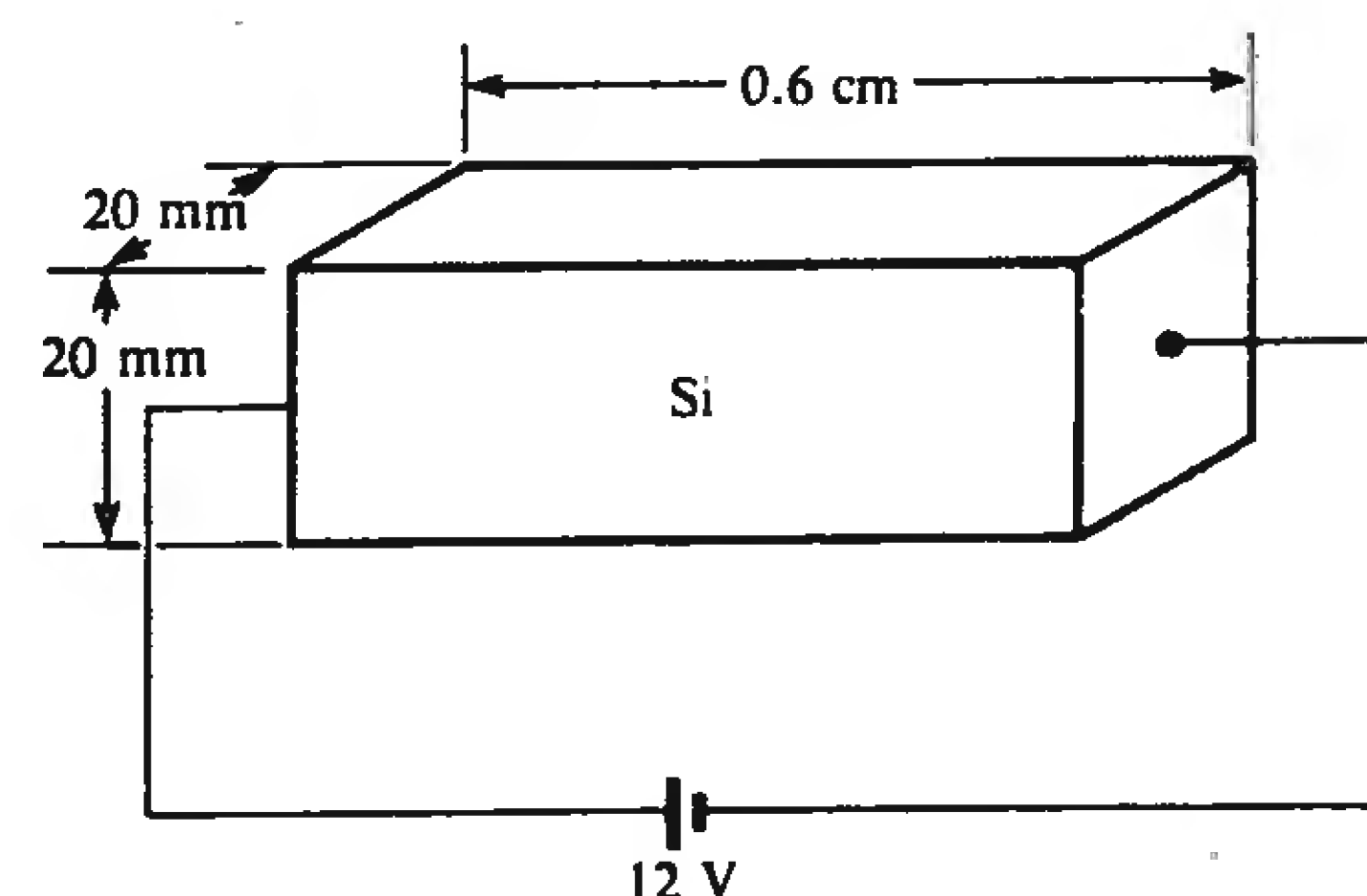
$$\left( \frac{\text{carriers}}{\text{m}^3} \right) \left( \frac{\text{coulombs}}{\text{carrier}} \right) \left( \frac{\text{V}}{\text{m}} \right) \left( \frac{\text{m}^2}{\text{V} \cdot \text{s}} \right) = \frac{\text{C}}{\text{s} \cdot \text{m}^2} = \frac{\text{A}}{\text{m}^2} = \text{current density}$$

#### EXAMPLE D-2

A potential difference of 12 V is applied across the ends of the intrinsic silicon bar shown in Figure D-5. Assuming that  $n_i = 1.5 \times 10^{16}$  electrons/m<sup>3</sup>,  $\mu_n = 0.14$  m<sup>2</sup>/(V · s), and  $\mu_p = 0.05$  m<sup>2</sup>/(V · s), find

1. the electron and hole velocities,
2. the electron and hole components of the current density,
3. the total current density, and
4. the total current in the bar.

FIGURE D-5  
(Example D-2)



#### Solution

We will assume that the electric field is established uniformly throughout the bar and that all current flow is along the horizontal axis of the bar (in the direction of the electric field).

1.  $\bar{E} = (12 \text{ V}) / (0.6 \times 10^{-2} \text{ m}) = 2 \times 10^3 \text{ V/m}$ . From equation D-3,

$$v_n = \bar{E} \mu_n = (2 \times 10^3 \text{ V/m}) [0.14 \text{ m}^2 / (\text{V} \cdot \text{s})] = 2.8 \times 10^2 \text{ m/s} \\ v_p = \bar{E} \mu_p = (2 \times 10^3 \text{ V/m}) [0.05 \text{ m}^2 / (\text{V} \cdot \text{s})] = 10^2 \text{ m/s}$$

2. Since the material is intrinsic,

$$p_i = n_i = 1.5 \times 10^{16} \text{ carriers/m}^3$$

and

$$J_n = n_i q_n v_n = (1.5 \times 10^{16}) (1.6 \times 10^{-19}) (2.8 \times 10^2) = 0.672 \text{ A/m}^2 \\ J_p = p_i q_p v_p = (1.5 \times 10^{16}) (1.6 \times 10^{-19}) (10^2) = 0.24 \text{ A/m}^2$$

3.  $J = J_n + J_p = 0.672 + 0.24 = 0.912 \text{ A/m}^2$

4. The cross-sectional area  $A$  of the bar is  $(20 \times 10^{-3} \text{ m}) \times (20 \times 10^{-3} \text{ m}) = 4 \times 10^{-4} \text{ m}^2$ . Therefore,  $I = JA = (0.912 \text{ A/m}^2)(4 \times 10^{-4} \text{ m}^2) = 0.365 \text{ mA}$ .

Recall that the resistance of any body can be calculated using

$$R = \frac{\rho l}{A} \quad (\text{D-5})$$

where

$R$  = resistance, ohms ( $\Omega$ )  
 $\rho$  = resistivity of the material,  $\Omega \cdot \text{m}$   
 $l$  = length, m  
 $A$  = cross-sectional area, m<sup>2</sup>

Conductance, which has the units of siemens (S), is defined to be the reciprocal of resistance, and conductivity is the reciprocal of resistivity:

$$\sigma = \frac{1}{\rho} \quad (\text{D-6})$$

Thus, the units of conductivity are  $1/(\Omega \cdot \text{m})$ , or siemens/meter (S/m).

The conductivity of a semiconductor can be computed using

$$\sigma = n \mu_n q_n + p \mu_p q_p \quad (\text{D-7})$$

Note that it is again possible to identify a component of conductivity due to electrons and a component due to holes.

#### EXAMPLE D-3

1. Compute the conductivity and resistivity of the bar of intrinsic silicon in Example D-2 (Figure D-5).
2. Use the results of (1) to find the current in the bar when the 12-V potential is applied to it.

#### Solution

1.  $n = p = n_i = p_i = 1.5 \times 10^{16} / \text{m}^3$ . From equation D-7,

$$\sigma = (1.5 \times 10^{16}) (0.14) (1.6 \times 10^{-19}) \\ + (1.5 \times 10^{16}) (0.05) (1.6 \times 10^{-19}) \\ = 4.56 \times 10^{-4} \text{ S/m}$$

Then  $\rho = 1/\sigma = 1/(4.56 \times 10^{-4}) = 2192.98 \Omega \cdot \text{m}$ .

2.  $R = \rho l / A = (2192.98) (0.6 \times 10^{-2}) / (4 \times 10^{-4}) = 32.89 \text{ k}\Omega$   
 $I = E / R = 12 / (32.89 \times 10^3) = 0.365 \text{ mA}$

The current computed this way is the same as that computed in Example D-2.



### Diffusion Current

Under certain circumstances, another kind of current besides drift current can exist in a semiconductor. Whenever there is a concentration of carriers (electrons or holes) in one region of a semiconductor and a scarcity in another, the carriers in the high-density region will migrate toward the low-density region, until their distribution becomes more or less uniform. In other words, there is a natural tendency for energetic carriers to disperse themselves to achieve a uniform concentration. Visualize, in an analogous situation, a small room crowded with people who are constantly squirming, crowding, and elbowing each other; if the room were suddenly to expand to twice its size, the occupants would tend to nudge and push each other outward as required to fill the new space. Another example of this natural expansion of energetic bodies is the phenomenon observed when a fixed quantity of gas is injected into an empty vessel: The molecules disperse to fill the confines of the container.

During the time that carriers are migrating from the region of high concentration to the one of low concentration, there is a transfer of charge taking place, and therefore an electric current. This current is called *diffusion current*, and the carriers are said to diffuse from one region to another. Diffusion is a transient (short-lived) process unless the region containing the higher concentration of charge is continually replenished. In many practical applications that we will study later, carrier replenishment does occur and diffusion current is thereby sustained.

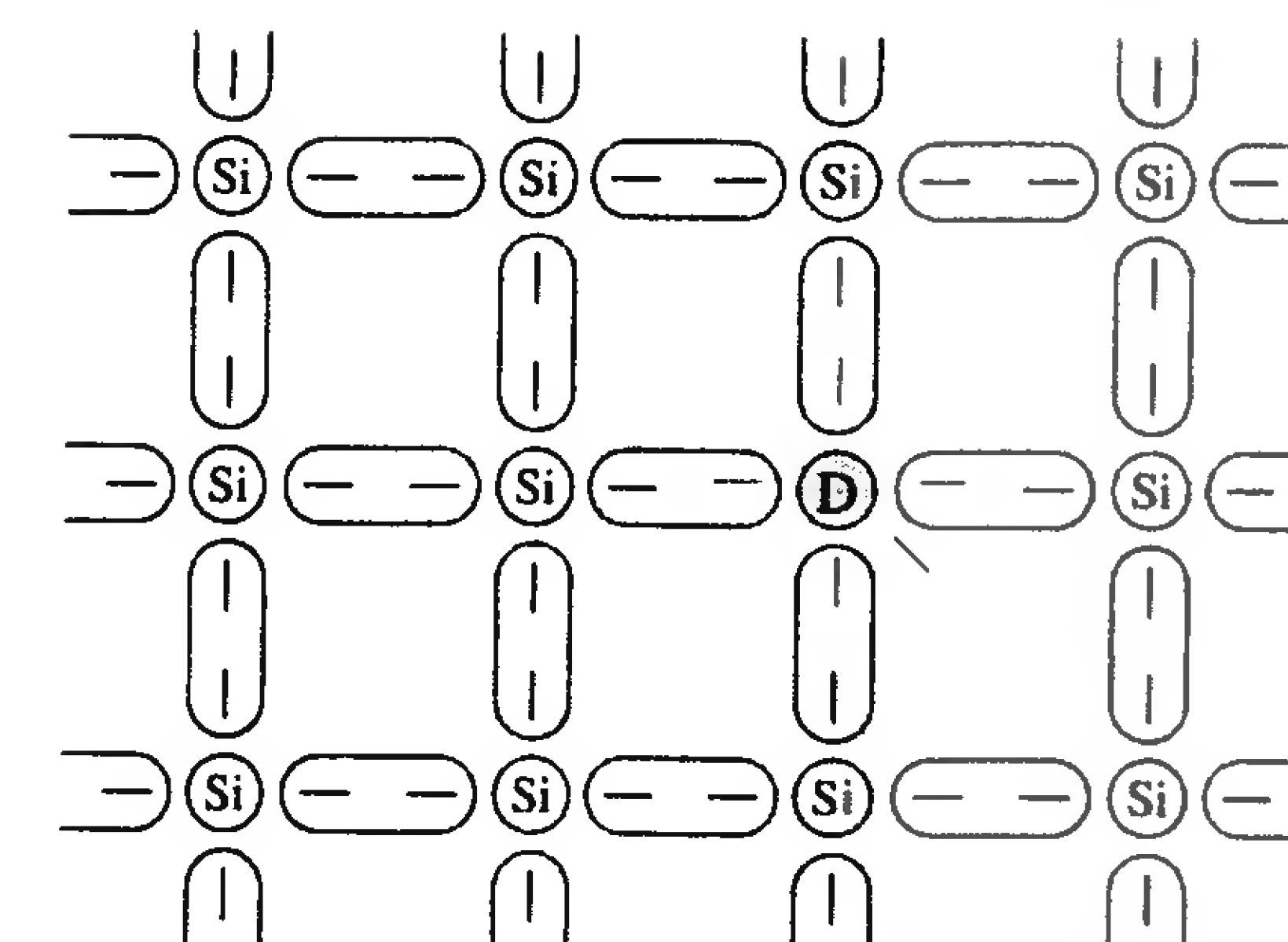
### D-5 *p*- AND *n*-TYPE SEMICONDUCTORS

Recall that intrinsic semiconductor material has the same electron density as hole density:  $n_i = p_i$ . In the fabrication of semiconductor materials used in practical applications, this balance between carrier densities is intentionally altered to produce materials in which the number of electrons is greater than the number of holes, or in which the number of holes is greater than the number of electrons. Such materials are called *extrinsic* (or *impure*) semiconductors. They are called *impure* because, as we will discuss presently, the desired imbalance is achieved by introducing certain impurity atoms into the crystal structure. Materials in which electrons predominate are called *n-type* materials, and those in which holes predominate are called *p-type* materials.

Let us first consider how *n-type* material is produced. Suppose that we are able, by some means, to insert into the crystal structure of a semiconductor an atom that has 5 instead of 4 electrons in its valence shell. Then 4 of those 5 electrons can (and do) participate in the same kind of covalent bonding that holds all the other atoms together. In this way, the impurity atom becomes an integral part of the structure, but it differs from the other atoms in that it has one "excess" electron. Its fifth valence electron is not needed for any covalent bond. Figure D-6 illustrates how the structure of a silicon crystal is modified by the presence of one such impurity atom. An impurity atom that produces an excess electron in this way is called a *donor* atom, because it donates an electron to the material. The nucleus of the donor atom is labeled D in Figure D-6. When a large number of donor atoms are introduced into the material, a correspondingly large number of excess electrons are created. Materials used as donor impurities in silicon include *antimony*, *arsenic*, and *phosphorus*.

Extrinsic semiconductor material is said to have been *doped* with impurity atoms, and the process is called *doping*. The impurity material is called a

FIGURE D-6 Structure of a silicon crystal containing a donor atom. The donor's nucleus is labeled D and the nuclei of the silicon atoms are labeled Si. The donor electrons are shown by colored dashes. Note the excess electron.

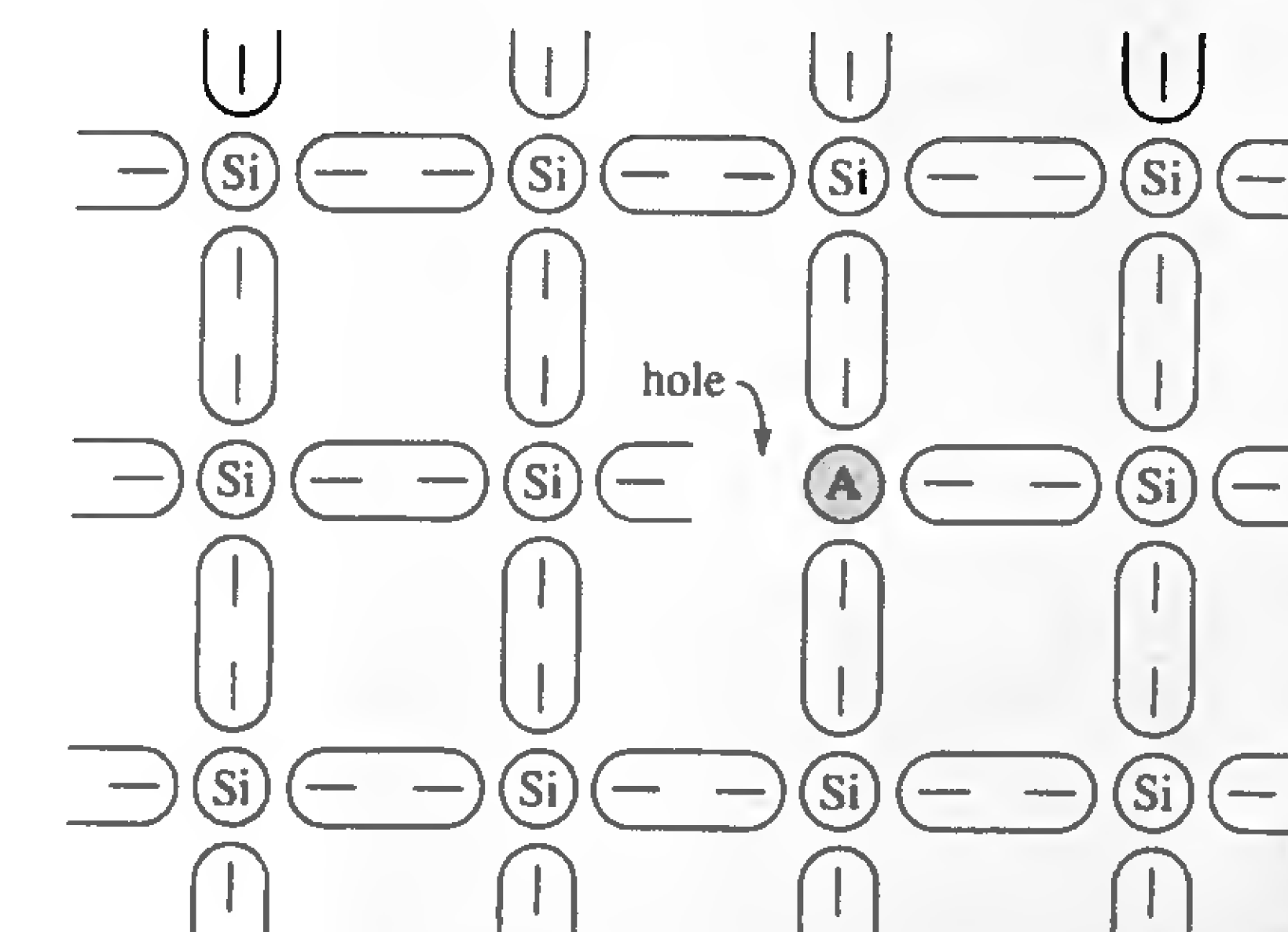


*dopant*. Because the silicon material illustrated in Figure D-6 has been doped with donor atoms and therefore contains an excess of electrons, it now constitutes *n-type* material. Note that the electrons are in "excess" only in the sense that there are now more electrons than holes; the material is still electrically neutral, because the number of protons in each donor nucleus still equals the total number of electrons that the donor atom brought to the material. Each excess electron is, however, very loosely bound to a donor atom and is, for all practical purposes, in the conduction band. The donor atom itself therefore becomes a positive ion. In all subsequent computations, we will assume that all impurity atoms are thus ionized.

*p-type* material is produced by doping a semiconductor with impurity atoms that have only three electrons in their outermost shells. When this kind of impurity atom is introduced into the crystal structure, an electron deficiency results, because the impurity atom contributes only three of the required four electrons necessary for covalent bonding. In other words, a hole is created everywhere the impurity atom appears in the crystal. Such impurity atoms are called *acceptors*, because the holes they produce can readily accept electrons. Figure D-7 shows a single acceptor atom in a silicon crystal. Materials used for doping silicon to create *p-type* material include *aluminum*, *boron*, *gallium*, and *indium*. Note once again that *p-type* material, like *n-type*, is electrically neutral because the electron deficiency exists only in the sense that there are insufficient electrons to complete all covalent bonds.

Although electrons are more numerous than holes in *n-type* material, there are still a certain number of holes present. The extent to which electrons dominate depends on the level of the doping: The more heavily the material is doped with donor atoms, the greater the degree to which the

FIGURE D-7 Structure of a silicon crystal containing an acceptor atom. The acceptor's nucleus is labeled A, and the nuclei of the silicon atoms are labeled Si. The acceptor electrons are shown by colored dashes. Note the incomplete bond and resulting hole caused by the acceptor's presence.





number of electrons exceeds the number of holes. In  $n$ -type material, electrons are said to be the *majority* carriers and holes the *minority* carriers. Similarly, the degree of acceptor doping controls the number of holes in  $p$ -type material. In this case, holes are the majority carriers and electrons the minority carriers.

An important relationship between the electron and hole densities in most practical semiconductor materials is given by

$$np = n_i^2 \quad (\text{D-8})$$

where

$n$  = electron density

$p$  = hole density

$n_i$  = intrinsic electron density

Equation D-8 states that the product of electron and hole densities equals the square of the intrinsic electron density. Because  $n_i = p_i$ , equation D-8 is, of course, equivalent to  $np = p_i^2 = n_i p_i$ . All of the theory we have discussed so far in connection with mobility, conductivity, and current density is applicable to extrinsic as well as intrinsic semiconductors. The carrier densities used in the computations are often found using equation D-8.

#### EXAMPLE D-4

A bar of silicon with intrinsic electron density  $1.4 \times 10^{16}$  electrons/m<sup>3</sup> is doped with impurity atoms until the hole density is  $8.5 \times 10^{21}$  holes/m<sup>3</sup>. The mobilities of the electrons and holes are  $\mu_n = 0.14$  m<sup>2</sup>/(V · s) and  $\mu_p = 0.05$  m<sup>2</sup>/(V · s).

1. Find the electron density of the extrinsic material.
2. Is the extrinsic material  $n$ -type or  $p$ -type?
3. Find the extrinsic conductivity.

**Solution**

1. From equation D-8,

$$n = \frac{n_i^2}{p} = \frac{(1.4 \times 10^{16})^2}{8.5 \times 10^{21}} = 2.3 \times 10^{10} \text{ electrons/m}^3$$

2. Since  $p > n$ , the material is  $p$ -type.

3. From equation D-7,

$$\begin{aligned} \sigma &= n\mu_n q_n + p\mu_p q_p \\ &= (2.3 \times 10^{10})(0.14)(1.6 \times 10^{-19}) + (8.5 \times 10^{21})(0.05)(1.6 \times 10^{-19}) \\ &= 5.152 \times 10^{-10} + 68 \approx 68 \text{ S/m} \end{aligned}$$

Note in the preceding example that the conductivity, 68 S/m, is for all practical purposes determined exclusively by the component of the conductivity due to holes, which are the majority carriers in this case. In practice, this is almost always the case: The conductivity essentially depends only on the majority carrier density. This result is due to a phenomenon called *minority carrier suppression*. To illustrate, suppose that the majority carriers are electrons, and that there are substantially more electrons than holes. Under these conditions, there is an increased probability that an electron-hole recombination (annihilation) will occur, thus eliminating both a free electron and a hole. Since there are very many more electrons than holes, the resultant percent decrease in electrons is much less than the percent decrease in holes. This effective suppression of minority carriers is

reflected in equation D-8 and leads to the following approximations, valid in most practical cases, for computing conductivity:

$$\begin{aligned} \sigma &\approx n\mu_n q_n \quad (n\text{-type material}) \\ \sigma &\approx p\mu_p q_p \quad (p\text{-type material}) \end{aligned} \quad (\text{D-9})$$

When one carrier type has a substantial majority, it is apparent from these equations that the conductivity of a semiconductor increases in direct proportion to the degree of doping with impurity atoms that produce the majority carriers.

#### D-6 THE $pn$ JUNCTION

When a block of  $p$ -type material is joined to a block of  $n$ -type material, a very useful structure results. The region where the two materials are joined is called a *pn junction* and is a fundamental component of many electronic devices, including transistors. The junction is not formed by simply placing the two materials adjacent to each other, but rather through a manufacturing process that creates a transition from  $p$  to  $n$  within a single crystal. Nevertheless, it is instructive to view the formation of the junction in terms of the charge redistribution that would occur if two dissimilar materials were, in fact, suddenly brought into very close contact with each other.

Let us suppose that a block of  $p$ -type material on the left is suddenly joined to a block of  $n$ -type material on the right, as illustrated in Figure D-8(a). In the figure, the acceptor atoms and their associated “excess” holes are shown in the  $p$  region. Remember that the  $p$  region is initially neutral because each acceptor atom has the same number of electrons as protons. Similarly, the donor atoms are shown with their associated “excess” electrons in the  $n$  region, which is likewise electrically neutral. Remember also that *diffusion* current flows whenever there is a surplus of carriers in one region and a corresponding lack of carriers of the same kind in another region. Consequently, at the instant the  $p$  and  $n$  blocks are joined, electrons from the  $n$  region diffuse into the  $p$  region, and holes from the  $p$  region diffuse into the  $n$  region. (Recall that this hole current is actually the repositioning of holes due to the motion of *valence-band* electrons.)

For each electron that leaves the  $n$  region to cross the junction into the  $p$  region, a donor atom that now has a net positive charge is left behind. Similarly, for each hole that leaves the  $p$  region (that is, for each acceptor atom that captures an electron), an acceptor atom acquires a net negative charge. The upshot of this process is that negatively charged acceptor atoms begin to line the region of the junction just inside the  $p$  block, and positively charged donor atoms accumulate just inside the  $n$  region. This charge distribution is illustrated in Figure D-8(b) and is often called *space charge*.

**FIGURE D-8** Formation of a  $pn$  junction. A = acceptor atom; h = associated hole; D = donor atom; e = associated electron; + = positively charged ion; - = negatively charged ion.

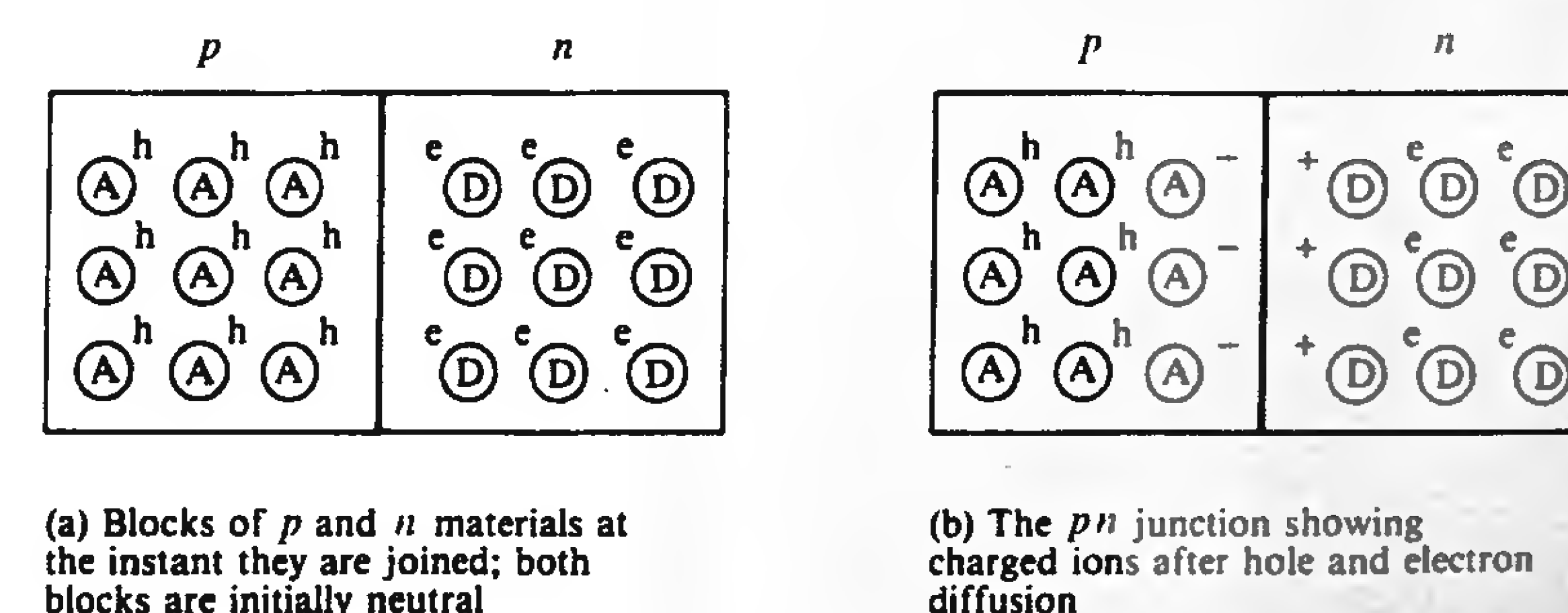
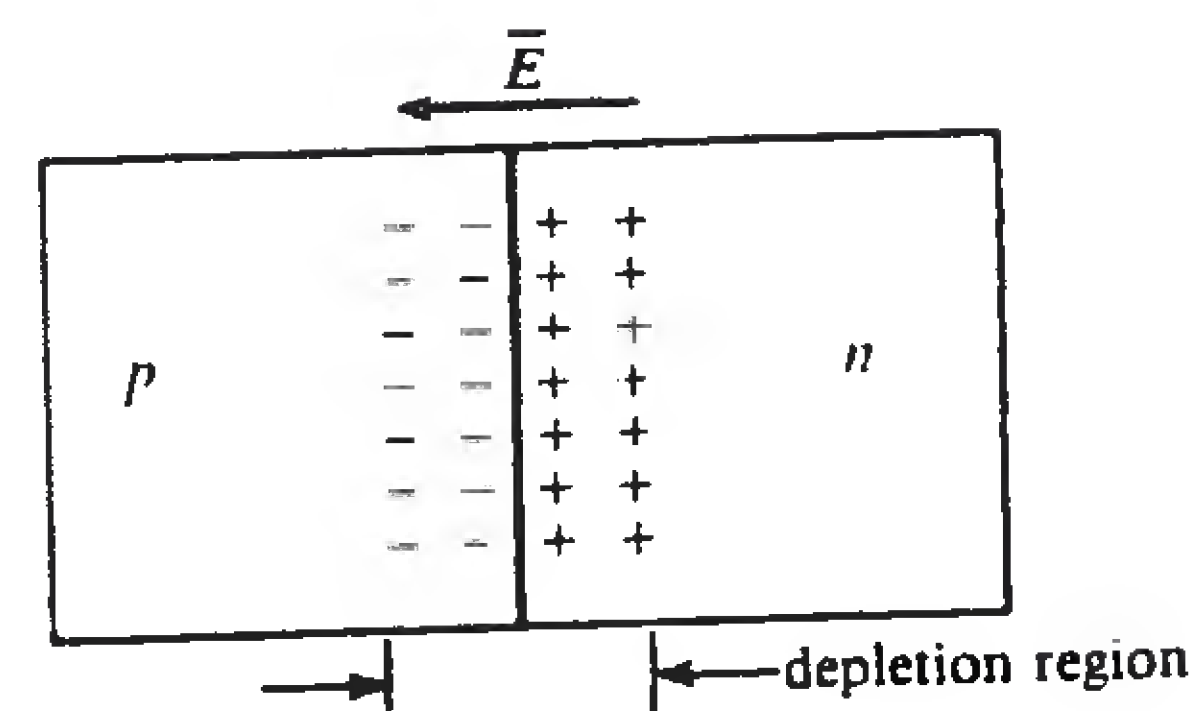




FIGURE D-9 The electric field  $\bar{E}$  across a  $pn$  junction inhibits diffusion current from the  $n$  to the  $p$  side. There are no mobile charge carriers in the depletion region (whose width is proportionally much smaller than that shown)



It is well known that accumulations of electric charge of opposite polarities in two separated regions cause an electric field to be established between those regions. In the case of the  $pn$  junction, the positive ions in the  $n$  material and the negative ions in the  $p$  material constitute such accumulations of charge, and an electric field is therefore established. The direction of the field (which by convention is the direction of the force on a positive charge placed in the field) is from the positive  $n$  region to the negative  $p$  region. Figure D-9 illustrates the field  $\bar{E}$  developed across a  $pn$  junction.

Note that the direction of the field is such that it *opposes* the flow of electrons from the  $n$  region into the  $p$  region, and the flow of holes from the  $p$  region into the  $n$  region. In other words, the positive and negative charges whose locations were *caused* by the original diffusion current across the junction are now inhibiting the further flow of current across the junction. An equivalent interpretation is that the accumulation of negative charge in the  $p$  region prevents additional negative charge from entering that region (like charges repel), and, similarly, the positively charged  $n$  region repels additional positive charge. Therefore, after the initial surge of charge across the junction, the diffusion current dwindles to a negligible amount.

The direction of the electric field across the  $pn$  junction enables the flow of *drift* current from the  $p$  to the  $n$  region, that is, the flow of electrons from left to right and of holes from right to left, in Figure D-9. There is therefore a small drift of *minority* carriers (electrons in the  $p$  material and holes in the  $n$  material) in the opposite direction from the diffusion current. This drift current is called *reverse* current, and when *equilibrium* conditions have been established, the small reverse drift current exactly cancels the diffusion current from  $n$  to  $p$ . The net current across the junction is therefore 0.

In the region of the junction where the charged atoms are located, there are no mobile carriers (except those that get swept immediately to the opposite side). Remember that the  $p$ -region holes have been annihilated by electrons, and the  $n$ -region electrons have migrated to the  $p$  side. Because all charge carriers have been depleted (removed) from this region, it is called the *depletion* region. See Figure D-9. It is also called the *barrier* region because the electric field therein acts as a barrier to further diffusion current, as we have already described. The width of the depletion region depends on how heavily the  $p$  and  $n$  materials have been doped. If both sides have been doped to have the same impurity densities (not the usual case), then the depletion region will extend an equal distance into both the  $p$  and  $n$  sides. If the doping levels are not equal, the depletion region will extend farther into the side having the smaller impurity concentration. The width of a typical depletion region is on the order of  $10^{-6}$  m. In the practical  $pn$  junction, there is not necessarily the abrupt transition from  $p$ - to  $n$ -type material shown in Figure D-8. The junction may actually be formed, for example, by a gradual increase in

the donor doping level of one block of  $p$ -type material, so that it gradually changes its nature from  $p$ -type to  $n$ -type with increasing distance through the block.

The electric field shown in Figure D-9 is the result of the *potential difference* that exists across the junction due to the oppositely charged sides of the junction. This potential is called the *barrier* potential because it acts as a barrier to diffusion current. (It is also called a *junction* potential, or *diffusion* potential.) The value of the barrier potential,  $V_0$ , depends on the doping levels in the  $p$  and  $n$  regions, the type of material (Si or Ge), and the temperature. Equation D-10 shows how these variables affect  $V_0$ :

$$V_0 = \frac{kT}{q} \ln \left( \frac{N_A N_D}{n_i^2} \right) \quad (\text{D-10})$$

where  $V_0$  = barrier potential, volts

$k$  = Boltzmann's constant =  $1.38 \times 10^{-23}$  J/K

$T$  = temperature of the material in kelvin ( $K = 273 + ^\circ\text{C}$ ; note that the correct SI unit of temperature is kelvin, *not*  $^\circ\text{K}$  or degrees Kelvin.)

$q$  = electron charge =  $1.6 \times 10^{-19}$  C

$N_A$  = acceptor doping density in the  $p$  material

$N_D$  = donor doping density in the  $n$  material

$n_i$  = intrinsic electron density

Note that the barrier potential is directly proportionate to *temperature*. As we shall see throughout the remainder of our study of semiconductor devices, temperature plays a very important role in determining device characteristics and therefore has an important bearing on circuit design techniques. The quantity  $kT/q$  in equation D-10 has the units of volts and is called the *thermal voltage*,  $V_T$ :

$$V_T = \frac{kT}{q} \text{ volts} \quad (\text{D-11})$$

Substituting equation D-11 into equation D-10,

$$V_0 = V_T \ln \left( \frac{N_A N_D}{n_i^2} \right) \quad (\text{D-12})$$

#### EXAMPLE D-5

A silicon  $pn$  junction is formed from  $p$  material doped with  $10^{22}$  acceptors/ $\text{m}^3$  and  $n$  material doped with  $1.2 \times 10^{21}$  donors/ $\text{m}^3$ . Find the thermal voltage and barrier voltage at  $25^\circ\text{C}$ .

**Solution.**  $T = 273 + 25 = 298$  K. From equation D-11,

$$V_T = \frac{kT}{q} = \frac{(1.38 \times 10^{-23})(298)}{1.6 \times 10^{-19}} = 25.7 \text{ mV}$$

$$n_i^2 = (1.5 \times 10^{16})^2 = 2.25 \times 10^{32}$$

From equation D-12,

$$V_0 = V_T \ln \left( \frac{N_A N_D}{n_i^2} \right)$$

$$= 0.0257 \ln \left( \frac{10^{22} \times 1.2 \times 10^{21}}{2.25 \times 10^{32}} \right)$$

$$= (0.0257)(24.6998) = 0.635 \text{ V}$$



## D-7 FORWARD- AND REVERSE-BIASED JUNCTIONS

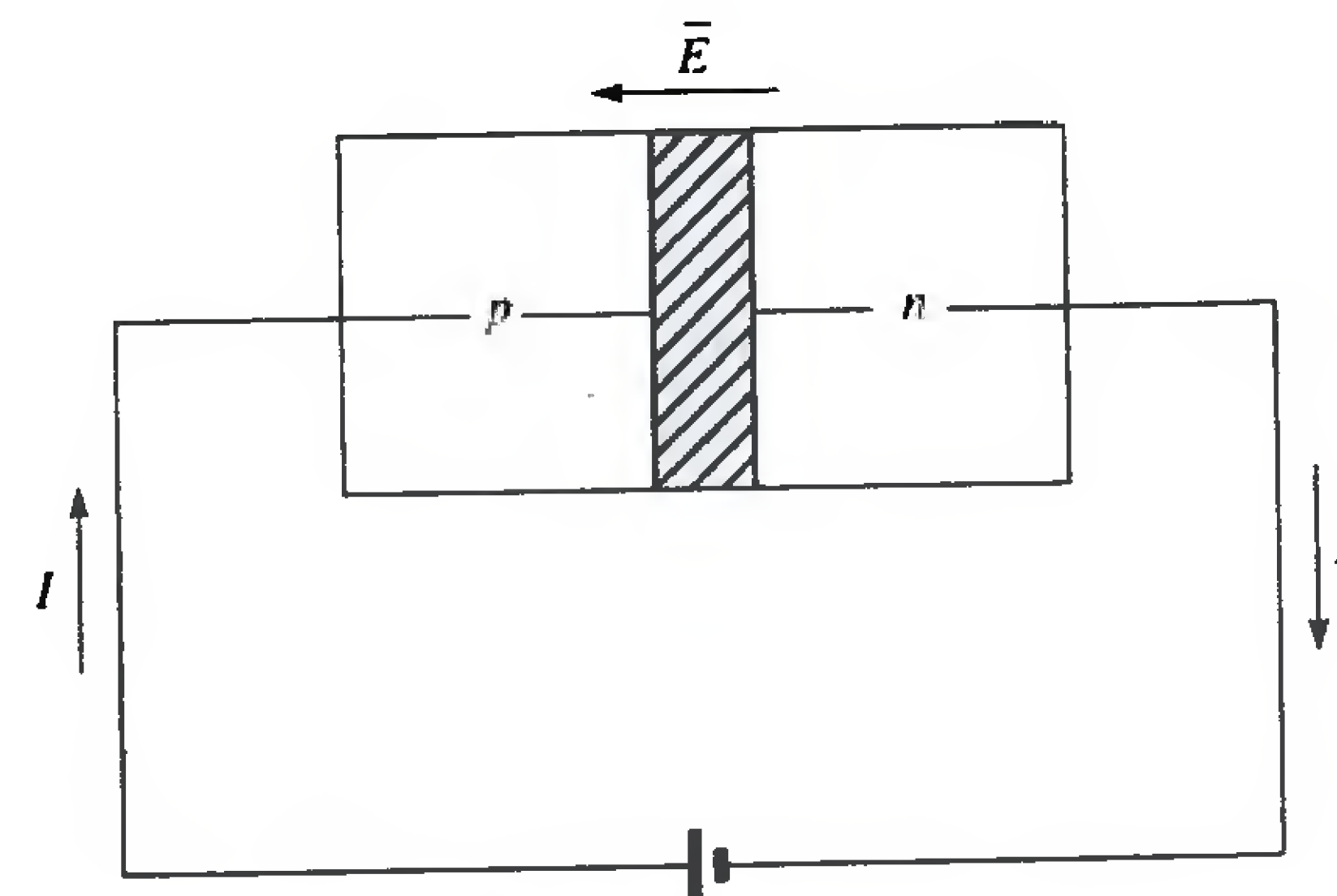
In the context of electronic circuit theory, the word *bias* refers to a dc voltage (or current) that is maintained in a device by some externally connected source. We discuss the concept of bias and its practical applications in considerable detail in Chapter 3. For now, suffice it to say that a *pn* junction can be biased by connecting a dc voltage source across its *p* and *n* sides.

Recall that the internal electric field established by the space charge across a junction acts as a *barrier* to the flow of diffusion current. When an external dc source is connected across a *pn* junction, the polarity of the connection can be such that it either opposes or reinforces the barrier. Suppose a voltage source *V* is connected as shown in Figure D-10, with its positive terminal attached to the *p* side of a *pn* junction and its negative terminal attached to the *n* side. With the polarity of the connections shown in the figure, the external source creates an electric field component across the junction whose direction *opposes* the internal field established by the space charge. In other words, the barrier is reduced, so diffusion current is enhanced. Therefore, current flows with relative ease through the junction, its direction being that of conventional current, from *p* to *n*, as shown in Figure D-10. With the polarity of the connections shown in the figure, the junction is said to be *forward biased*. (It is easy to remember that a junction is forward biased when the positive terminal of the external source is connected to the *p* side, and the negative terminal to the *n* side.)

When the *pn* junction is forward biased, electrons are forced into the *n* region by the external source, and holes are forced into the *p* region. As free electrons move toward the junction through the *n* material, a corresponding number of holes progresses through the *p* material. Thus, current in each region is the result of majority carrier flow. Electrons diffuse through the depletion region and recombine with holes in the *p* material. For each hole that recombines with an electron, an electron from a covalent bond leaves the *p* region and enters the positive terminal of the external source, thus maintaining the equality of current entering and leaving the source.

Since there is a reduction in the electric field barrier at the forward-biased junction, there is a corresponding reduction in the quantity of ionized acceptor and donor atoms required to maintain the field. As a result, the depletion region *narrows* under forward bias. It might be supposed that the forward-biasing voltage *V* could be increased to the point that the barrier field would be completely overcome, and in fact reversed in direction. This is not, however, the case. As the forward-biasing voltage is increased, the corresponding increase in current causes a larger voltage drop across the *p* and *n* material outside the depletion region, and the barrier field can never shrink to 0.

FIGURE D-10 A voltage source *V* connected to forward bias a *pn* junction. The depletion region (shown shaded) is narrowed.



## D-8 A SEMICONDUCTOR GLOSSARY

- Acceptor** An impurity atom used in the doping process to create a hole in a semiconductor crystal; contains 3 electrons in its outermost valence shell.
- Annihilation** See *recombination*.
- Anode** The *p* side of a *pn* junction diode.
- Avalanching** Large current flow through a reverse-biased diode when it breaks down; caused by a high electric field imparting high velocities to electrons that then rupture covalent bonds.
- Band, energy** See *energy band*.
- Barrier diode** See *Schottky diode*.
- Barrier potential** Potential (voltage) established by the presence of layers of charge lying on opposite sides of a *pn* junction.
- Barrier voltage** See *barrier potential*.
- Bias** Connection of an external voltage source across a *pn* junction. See also *forward bias* and *reverse bias*.
- Boltzmann's constant (*k*)** Constant used in the diode equation;  $k = 1.38 \times 10^{-23}$  J/K.
- Breakdown voltage** The reverse-biasing voltage across a diode that causes it to conduct heavily in the reverse direction (cathode to anode).
- Carrier** See *charge carrier*.
- Carrier density** The number of carriers (holes or electrons) per cubic meter of semiconductor material. See also *electron density* and *hole density*.
- Cathode** The *n* side of a *pn* junction diode.
- Charge carrier** An electron, which carries one unit of negative charge, *q*, or a hole, which carries one unit of positive charge.
- Charge of an electron (*q*)**  $q = 1.6 \times 10^{-19}$  coulombs (C).
- Conduction band** The energy band of free electrons.
- Conductivity ( $\sigma$ )** A measure of the ability of a particular material to conduct current; units: S/m; the reciprocal of resistivity.
- Covalent bond** Shared electrons held to a parent atom in a crystal.
- Current density (*J*)** Current per unit cross-sectional area, A/m<sup>2</sup>.
- Depletion region** Region in a *pn* junction where there are no mobile charge carriers.
- Diffusion current** Migration of carriers from a region where there are a large number of their own type to a region where there are fewer.
- Diffusion potential** See *barrier potential*.
- Diode** A *pn* junction. A discrete diode is a *pn* junction fitted with an enclosure and terminals for connection to its *p* and *n* sides (anode and cathode).
- Donor** An impurity atom used in the doping process to create a free electron in a semiconductor crystal; contains 5 electrons in its outermost valence shell.
- Doping** The process of introducing impurity atoms into a crystal to create *p*- or *n*-type material.
- Drift current** Current created by the motion of charge carriers under the influence of an electric field.
- Electron, charge on (*q*)** See *charge of an electron*.
- Electron current** The flow of (free) electrons.
- Electron density (*n*)** The number of free electrons per cubic meter of material.
- Electron, free** See *free electron*.
- Electron-volt (eV)** Unit of energy; the energy acquired by an electron when accelerated through a potential difference of 1 V:  $1 \text{ eV} = 1.602 \times 10^{-19}$  J.
- Emission coefficient ( $\eta$ )** Coefficient used in the diode equation; its value depends on voltage and type of material;  $1 \leq \eta \leq 2$ .
- Energy band** A range of energies possessed by electrons. See also *conduction band*, *forbidden band*, and *valence band*.
- Energy gap** The width of the forbidden band; the difference in energy levels between the conduction and valence bands.